Hugh O'Connor

Using the Electronic Information Ecosystem for Research

Today there are many more sources for professional reference than there used to be. Many of these are entirely digital, without any print equivalent. In a sense, the Internet has introduced a "habitat" within which many new "species" of professional and popular reference tools have established a niche for themselves. This broadly increases the information available to knowledge workers, and also complicates the task of finding, evaluating and using all of it.

The major Web search services are dynamic in coverage and functions. Changes in site design and in the details of the services are made frequently, as each of the search engines, directories and subject guides compete to attract the traffic of the Net. Like the rest of the Internet, they evolve toward improvement.

What are some of these searching tools? Even though they are relatively rare compared with the totality of the Internet, there are still many of them from which to choose. One Web site (**World Wide Web Search Engines & Directories** at <http://www.lawresearch.com/cewwwd.htm>) lists nearly a thousand such searching services. The vast majority cost nothing to use.

In this discussion we are not talking about the classic fee-based database searching services that have been around, under one name or another, since the early 1970s. Although they continue to play a significant role in the information industry, their coverage of information on the Internet at large is still quite limited. They still index or contain the full-text of the premium, often print-equivalent information sources (newspapers, journals, and specialized data collections) that have been their specialty since their inception. They have, however, largely adopted the Web as the channel of access through which searchers now connect to those databases. A useful page of links to the Web incarnations of the major database search services can be found at On-line Inc.'s **SuperSites** page at <http://www.onlineinc.com/corporate/supersites.html>.

There are a number of Web-based search engines, using computer-based indexing methods, that attempt to offer very broad, if not comprehensive, coverage of the content of the World Wide Web. The most well-known and well-regarded of them (in alphabetical order) are:

Alta Vista <http://www.altavista.digital.com>
Excite <http://www.excite.com>
HotBot <http://www.hotbot.com>).
Infoseek <http://www.infoseek.com>).
Lycos <http://www.lycos.com>).
Northern Lights <http://www.nlsearch.com>
Open Text Index <http://index.opentext.net>
WebCrawler <http://www.webcrawler.com>

These services are extremely popular, but require some study to use them to best effect. Because they index words within some or all of the texts of the hundreds of thousands or millions of Web pages they index, they tend to retrieve hundreds and sometimes thousands of Web pages in response to the simplest search.

Each of them has some form of "relevance ranking" built into its search method, so that the pages identified as most likely to be relevant to a search are presented first. This is done to save a searcher's time. The expectation is that, as you go down a list of perhaps thousands of entries, their appropriateness and usefulness to you will drop off more or less rapidly, and you will soon reach a point below which the retrieved items are no longer likely to be of interest to you.

Although all the search engines listed above index the World Wide Web, they go about their work in different ways. Some index the words at a Web site's home page, and others go deeper than that. Some search engines cover a more selective set of Web sites than others. Also, the search rules are somewhat different from one engine to another. Since the search engines cover overlapping but not identical material; and since they operate differently, you will get different results from different search engines, at different times. Also, a lot of new material is always appearing on the Web, while the existing sites are always changing and being updated. The consequence is that you should be prepared to frequently re-run a search on any topic in which you are strongly interested—new material goes online every day.

Even the best indexes of Internet content are substantially behind the actual changes that go on constantly over the Net. Search engines work on the basis of indexes that are created by automated surveys of the state of the Web. The indexes have to be frequently rebuilt, since many of the Web's details will change sooner or later after they have been surveyed and recorded. Also, much new

material appears on the Web every day. One commentator has likened the task of indexing the World Wide Web to trying to paint an unfinished bridge that gets larger as you paint it.

This is not the only limitation on search engines' accuracy and range of coverage. Some Web sites, at their Webmasters' request and for whatever reason, are excluded from search engine indexes. Other Web sites contain part of their content in formats that are technically not indexable by the search engines.

Sometimes a Web page is not available even though it turns up in your search results. There are many possible explanations for this. The indexed site may be "off-line" for routine maintenance or because of a system problem, and will be available again soon. Popular sites get overwhelmed and may respond with a "busy signal." The indexed Web page may no longer be maintained at the address the search engine indicated. The page may have been moved, given a new file name, or removed altogether. This could happen at any time, because the maintainers of Web and other Internet sites have complete control over whether or not and how long they will keep their site available. In a sense, the price of the Net's attractive features—its creativity, spontaneity and decentralization—is that it changes so often as to be occasionally unreliable.

There are a number of Web sites dedicated to tracking and evaluating the major Web search engines. These sites are particularly useful because of their comparative approach. Some are addressed more to archivists and librarians and other professional researchers, and provide the results of in-depth performance tests of the search engines. Others are addressed more to Internet users in general, and are more succinct and less academic in their approach. All of them offer perspective on the search engines and advice on their use.

Search Engine Watch: News, Tips, and More About Search Engines <www.searchenginewatch.com>. This is probably the most comprehensive of the sites, and has information of interest to Web professionals as well as general readers. The section called "Search Engine Facts and Fun" is the part of the site most appropriate for novices and general Internet users, and contains information on how to use the search engines to best effect. A free monthly email newsletter, called the *Search Engine Report*, is offered to those who register with this site.

PC Magazine's Complete Guide to Searching the Net <www.zdnet.com/pcmag/features/websearch/_open.htm>. This is the online version of a feature article for the December 2, 1997 issue of the magazine. It consists of evaluative comments on each of the search engines, as well as some of the metasearch services. It is addressed to the general Internet user.

C/Net's Search Engine Shoot-Out: Top Engines Compared <www.cnet.com/Content/Reviews/Compare/Search2/?st.cn.fd.ccol.re>. Like **PC Magazine's Guide**, above, this is a recent (January 1998) comparative evaluation of search engines and metasearch services, aimed at the general Internet user. Note especially that this feature article includes a page from which you can test the major browsers for yourself.

The Search is Over: Search Engine Secrets of the Pros <http://www.zdnet.com/pccomp/features/fealO96/sub2.html>. This 1997 (otherwise undated) review of the search engines offers very specific tips and advice for using each of the major services. Since the search engine details may change as they are further developed, this article needs to be read circumspectly. If any of its tips do not work for you, this may be an indication of technical change in the search engines themselves.

Search Engine Showdown: Comparing Internet Finding Tools <http://www.imt.net/~notess/search/about.html). This excellent site is the product of Greg Notess, a librarian and author of many articles about online information searching. In his words, "This site summarizes, reviews, and compares the search features and database scope of the Internet search engines and finding aids." It is addressed primarily to other librarians and professional information searchers, and it contains a selected bibliography of print and online sources (with links) for those who want to become experts on the subject of search engines.

Because the retrieval results of the major search engines are at best statistical samplings of the Web's content, some other types of Web tools have an important role in Internet research. A few of these include:

ProFusion <http://www.designlab.ukans.edu/profusion/>. ProFusion is one of the so-called "metasearch" services, which deal with the variability in the search results of each of the major search engines by searching them all, and collating the results. This especially "intelligent" metasearch service will put your search terms into the various formats used by the search engines it employs, will eliminate duplicate citations retrieved by different engines, and will relevance-rank the search results from different engines into a single sequence. ProFusion offers a personalized search service, allowing users to register their regular queries. Those queries are periodically re-run, and the system tells you when there are new results.

The Argus Clearinghouse <http://www.clearinghouse.net>, **The World Wide Web Virtual Library** <http://www.w3.org/vl/>, **InfoMine**

<http://lib-www.ucr.edu/> and the **Virtual City of Alexandria** <http://www.alexandria-home.com/> are all attempts to organize the content of the World Wide Web into the subject arrangement of a library. Web sites like these are often referred to as "metasites." You choose your broad subject of interest from a list and these services connect you with a detailed page full of hyperlinks to the Internet resources offering information on your chosen topic. Coverage in these services is selective, but still very extensive. Because human judgment is involved, they offer greater precision—although perhaps not as much recall—than the search engines.

The Internet Sleuth <http://www.isleuth.com>. The Sleuth connects with "over 2,000 searchable databases" on the Web. It organizes those databases into broad categories like "News," "Sciences,," and "Travel" and permits you to search up to six databases in any given category at a time. Click the "About" button on the home page for a short, useful discussion on how to search the Internet Sleuth to best effect. Such gateways to Internet databases are particularly significant because search engines themselves do not usually index database contents.

Intemets <http://www.internets.com>. This is another very large collection of links to databases, organized quite a bit like *The Internet Sleuth*. This service adds a current newsfeed and a planetary weather map to its home page. It claims to be "the biggest filtered collection of useful search engines anywhere on the World Wide Web."

Professional reference tools now include software. Along with the proliferation of Web sites offering guidance to the scattered resources of the Internet, there has been a parallel development of software products specifically geared to assisting in the process of Internet research. Some such programs include utilities that, for instance, help to manage and organize Web bookmark lists. Others are more elaborate, serving as "intelligent agents" that search, compare, analyze, compile and organize information from many Internet sources while their human client is doing something else. We are at a point in the flowering of the software industry in which many products are available—many inexpensive or free of charge—that help researchers to analyze Web sites, find new sites of interest, manipulate digital text, and do research generally in ways impossible only a few years ago.

Some Web sites have appeared that not only list programs in these categories, but serve as links to the sources of the software themselves. Although some of these programs are commercial products, many are freeware or at least shareware, and are downloadable. Some good lists to check include:

Web SearchUser Tools <http://www.zdnet.com/products/searchuser/tools.html>. This is a short listing of both major Web search sites and Web search software, with links to each.

Botspot <http://www.botspot.com>. This site is concerned specifically with intelligent agent and "bot" software.

Cool Tools(<http://www.cooltool.com>. This source is not exclusively concerned with software that affects Web and Internet searching. It is, rather, a good, general, critical review source on new Internet-related software in general.

TUCOWS <http://www.tucows.com>. TUCOWS (The Ultimate Compilation of Winsock Software) is one of the major compilations of Intemet-related software of any kind. Search-related software is only a small part of what it offers.

Some software sites that offer interesting products for Web research, navigation and data compilation include the following. This is not at all a comprehensive list, and some or all of these items may be found listed at one or more of the sites indicated directly above. Most of these programs are available either free or in a free version from their sponsoring Web sites.

Alexa <http://www.alexa.com>
Autonomy <http://www.agentware.com>
Citizenl <http://www.citizenl.com>
WebFerret <http://www.ferretsoft.com>
WebTurbo <http://www.webturbo.com>

Looking at the Mammoth bison corral at Yellowstone. Mounted ranger is visible inside the corral. Bunsen Peak is in the background. Photo courtesy Yellowstone NP Archives.

Hugh O'Connor is with the Research Information Center, American Association of Retired Persons. This article is adapted from his speech at the Information Ecosystem Conference.